M.A. Zernova, M.V. Tkacheva
Ural Federal University named after the first President of Russia B.N. Yeltsin
Yekaterinburg, Russia

METHODS TO PREVENT SPAM ATTACKS IN SOCIAL NETWORKS

Abstract: This paper deals with the phenomenon of spam in social networks and shows effective methods of preventing spam. In the first part, the concepts of spam, social networks are studied, and the history of the appearance of spam is described. The second part provides brief information about renaming a user account, account markets where the user account data is exchanged for a fee or other limited services. It is revealed that this operation and the use of the services of these markets can be a direct source of spam. In the third part, the methods of the authors of some articles are investigated. It is described in details how to deal with different types of spam. Three spam detection modules are analyzed: a COMPA system that helps to identify a compromised account; S3D – a spam detection module based on four light detectors; and FRAppE – a rigorous Facebook application evaluator. In addition, this part provides information on the results of various experiments conducted with these modules. Based on the studied information about modules and types of spam, the final part concludes on the effectiveness of their use and provides the results.

Keywords: spam, social network, name capture, account markets, spam detector.

М.А. Зернова, М. В. Ткачева
Уральский федеральный университет имени первого Президента России Б.Н. Ельцина
Екатеринбург, Россия

МЕТОДЫ ПРЕДОТВРАЩЕНИЯ СПАМ-АТАК В СОЦИАЛЬНЫХ СЕТЯХ

Аннотация: В данной работе исследован феномен спама в социальных сетях и изучены эффективные методы предотвращения спама в социальных сетях. В первой части изучены понятия спама, социальных сетей, а также исследована история появления спама. Во второй части дана краткая информация о переименовании аккаунта пользователя, а также о рынках аккаунтов, которые предоставляют подписчиков в обмен на плату или другие ограниченные услуги в обмен на данные учетной записи пользователя и выявлено, что данная операция и использование услуг таких рынков могут стать прямым источником спама, который будет сыпаться на пользователя, допустившего ошибку. В третьей части исследованы методики авторов некоторых статей, в которых подробно разобрано, как справляться с различными видами спама. Изучены три модуля обнаружения спама: COMPA – система, которая поможет определить скомпрометированный аккаунт; S3D – модуль обнаружения спама, основанных на четырех легких детекторах; и FRAppE – строгий оценщик приложений Facebook. Здесь же приведена информация о результатах различных экспериментов, проведённых с данными модулями. На основе изученных модулей и видов спама в заключительной части делается вывод об эффективности их использования и дается информация о полученных результатах.

Ключевые слова: спам, социальная сеть, захват имени, рынки аккаунтов, детектор спама.

It is difficult to imagine modern reality without social networks. A social network is a platform that is designed to provide relationships

between people or organizations on the Internet. This is a unique platform for communication, sharing news, getting information and just for relaxing. However, spam can block the user to get the desired function in social networks.

Spam is a message that is sent to people who did not consent to receive it. Obviously, it would be unpleasant to miss an important letter in a heap of spam. Usually such messages are sent for advertising, distribution of malicious programs and phishing. It is interesting to note that the word spam is derived from the name of canned meat, which was annoyingly advertised after the end of the Second World War.

Previously, spam was mainly distributed via email, as it was the main communication tool. It was easy to collect email addresses through chat rooms, websites, and customer lists. In the end, email filters have become more fastidious and more effective in terms of detecting spam emails. However, scammers have found a new target: social networks. So, spam appeared in social networks – a phenomenon that is characterized by hacking accounts or entering into the trust of the user to spread unwanted information.

The rationale of the problem of spam in social networks is beyond doubt. 40% of accounts in social networks are currently spam. Such fake accounts are the main key to the spread of spam in social networks. To gain trust, attackers try to become «friends» or join already verified accounts, for example, accounts of celebrities or public figures. If you join such an account, spam will start pouring on you. Another way to attack – hacking account, managing it by distributing fake messages from the user. So, you may receive a message with a link from your friend or unfamiliar person. When you walk through it out of curiosity or carelessness, you can get caught in the trick of the fraudster. Of course, this does not mean that it is necessary to «block» all the links that are sent to you, but you should pay attention to the content of the message, punctuation, grammar, because spam messages usually have problems with these aspects.

So, the problem of this work is the lack of effective ways to protect users from spam in social networks.

Purpose: to study effective methods of protection from spam of ordinary users.

1. Unusual type of spam in the network.

As it turned out, to become a victim of spam is quite simple. One wrong mouse click or one wrong link click as spam immediately begins to pour on you. However, if many people realize that clicking on unknown

links from unknown users is wrong, then some other types of spam are not available to everyone.

It would seem that an innocuous operation – changing the account name, but this is a great operation for intruders on the Internet. By freeing your «past» name, you give the attacker the chance to publish advertisement, share malicious links on your behalf.

Capture the name lies in the fact that the attackers set the profile names similar to names of popular accounts. Thus, they can attract a large number of victims to spread malicious links, aggressive content and advertising. Of the 10% of public tweets per month, 3% of the accounts selected are potentially malicious. There is only one way to combat this – to tighten the security rules on Twitter, including prohibiting the use of the account name [1].

There is another unusual type of spam. This spam through accounts, bought market accounts. These are markets that promise their subscribers to provide subscribers in exchange for a fee or limited services free, but in exchange for Twitter user account credentials. The services of such markets are often associated with abusive behavior and compromised Twitter profiles. There are some criteria by which you can identify accounts that have been purchased by the Twitter account market. This includes, for example, posting unrelated updates on trending topics or so-called spam mentions where a large number of tweets mention users that have no relation with the account that sends the tweets. The way these markets operate directly violates Twitter's terms of service [2].

2. Some real ways to protect against spam.

If you cannot cope with spam on your own, good helpers will be developed. One of these, for example, is a semi-supervised spam detection framework, named S3D.

S3D utilizes four lightweight detectors to detect spam tweets on real-time basis and update the models periodically in batch mode.

The spam detection module consists of four lightweight detectors:

1. Blacklisted domain detector labels tweets containing blacklisted URLs;
2. Near-duplicate detector labels tweets that are near-duplicates of confidently prelabeled tweets;
3. Reliable ham detector labels tweets that are posted by trusted users and that do not contain spammy words;
4. Multiclassifier-based detector labels the remaining tweets.

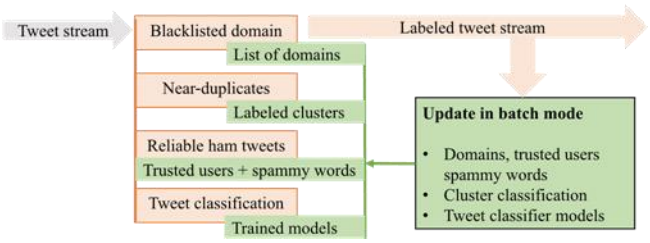

Figure 1 – System overview of the S^3D framework.

The information required by the detection module is updated in batch mode based on the tweets that are labeled in the previous time window. Experiments on a large-scale data set show that the framework adaptively learns patterns of new spam activities and maintain good accuracy for spam detection in a tweet stream [3].

This module will help the user to recognize spam even where it is not visible to the naked eye, but the creators of the COMPA are based on identifying accounts that are hijacked by attackers and spamming.

COMPA is the first detection system designed to identify compromised social network accounts. COMPA is based on a simple observation: social network users develop habits over time, and these habits are fairly stable. A typical social network user, for example, might consistently check the posts in the morning from the phone, and during the lunch break from the desktop computer. Furthermore, interaction will likely be limited to a moderate number of social network contacts. Conversely, if the account falls under the control of an adversary, the messages that the attacker sends will likely show anomalies compared to the typical behavior of the user. If the message does not match the normal user behavior, the system marks it as a possible attack by intruders [4].

There is also another way to deal with unwanted mailings – FRAppE. FRAppE (Facebook's Rigorous Application Evaluator) is the first tool focused on detecting malicious apps on Facebook. To develop FRAppE, the information is collected by observing the posting behavior of 111K Facebook apps seen across 2.2 million users on Facebook. First, it is analyzed a set of features that help to distinguish malicious and benign applications. For example, it is discovered that malicious apps often share common names with other apps and typically request fewer permissions than secure apps. Secondly, using these distinctive features, it is shown that FRAppE can detect malicious applications with an accuracy of 99.5%, without false positives and with a high true positive rate (95.9%). The Facebook ecosystem of malicious apps has also been studied and the mechanisms that these apps use for distribution have been identified.

Interestingly, many applications collude and support each other; in the dataset, 1,584 applications have been found allowing the viral spread of 3,723 other applications through their messages. In the long term, FRAppE is seen as a step towards creating an independent watchdog to evaluate and rank apps to alert Facebook users before installing apps [5].

Thus, using the above modules is very useful and effective for the average user in the network. The problem of this work is partially solved. This article will help users to find the difference between the types of spam, teach users not to succumb to malicious attacks, and use useful and necessary modules to combat spam.

REFERENCES

1. Ahmad S., Egele M., Maticonti E., Nikiforakis N.,Nikiforou N., Onaolapo J., Stringhini G. Why Allowing Profile Name Reuse Is A Bad Idea University College London, Stony Brook University, Boston University. – 2016, 6 pages.
2. Egele M., Kruegel C., Stringhini G., Vigna G. Poultry Markets: On the Underground Economy of Twitter Followers. University of California, Santa Barbara – 2012, 6 pages.
3. Sedhai S., Sun A. Semi-supervised Spam Detection in Twitter Stream. Vol. 5 of IEEE transactions on computational social systems. – 2018, 8 pages.
4. Egele M., Kruegel C., Stringhini G., Vigna G.Towards Detecting Compromised Accounts on Social Networks. Vol. 14 of IEEE transactions on dependable and secure computing. – 2017, 14 pages.
5. Faloutsos M., Huang T., Madhyastha H., Rahman S. Detecting Malicious Facebook Application. Vol. 24 of IEEE/ACM transactions on networking. – 2016, 17 pages.